RAW SEQUENCE LISTING   DATE: 05/08/2002
PATENT APPLICATION: US/09/720,384A   TIME: 17:30:19

Input Set : A:\BB-1167-B USPCT Corrected Seq List 02.txt
Output Set: N:\CRF3\05082002\I720384A.raw

```
 3 <110> APPLICANT: Falco, Saverio
 4        Allen, Stephen
 5        Anderson, Shawn
 7 <120> TITLE OF INVENTION: Genes Encoding Sulfate Assimilation Proteins
 9 <130> FILE REFERENCE: BB-1167-B
11 <140> CURRENT APPLICATION NUMBER: 09/720,384A
12 <141> CURRENT FILING DATE: 2000-12-21
14 <150> PRIOR APPLICATION NUMBER: 60/092,833
15 <151> PRIOR FILING DATE: 1998-07-14
17 <160> NUMBER OF SEQ ID NOS: 14
19 <170> SOFTWARE: Microsoft Office 97
21 <210> SEQ ID NO: 1
22 <211> LENGTH: 890
23 <212> TYPE: DNA
24 <213> ORGANISM: Zea mays
26 <400> SEQUENCE: 1
27 ggtcagcggc ggcggccgtc gcagggatca gcagcagcag cagcgcgctg gtgacctcga  60
28 ccgtcgggaa atcgacgaac atcctgtggc atgagtgcgc catcgggcag aaggagcgac 120
29 agggtctgct gaaccagaag ggctgcgtcg tgtggatcac tggcctaagc ggttcaggga 180
30 aaagcacgct cgcgtgcgcg ctgagccgcg agctgcacgg cagaggccac ctcacgtacg 240
31 tcctcgacgg cgacaacctc aggcacgggc tgaacaggga cctcagcttc ggagcagagg 300
32 accgcgccga gaacatccgc agagtagggg aagtagcgaa gctgttcgcc gacgctggcc 360
33 tcgtctgcat cgccagcctc atatcgccct acagaagcga ccgaagcgcg tgtcgcgatc 420
34 tgctgcccaa gcactcgttt atcgaggtgt cctggacgt gccgcttcaa gtgtgcgaag 480
35 ccagggaccc caaaggcctc tacaagctcg cacgcgccgg caaaatcaaa gggttcaccg 540
36 gcatcgacga tccttacgaa ccgccgtcgg actgtgagat agtgatccag tgtaaagtcg 600
37 gcgactgccc ttcgcctgaa tcgatggctg gtcacgttgt gtcgtacctt gagacgaatg 660
38 gtttcctcca ggactagaca tggaatgcga tcgatgcgtc tgatgtgtat atatgtagca 720
39 gcagccggag cggcattgcc aaggctgtgt aatctcatgg ctgtctttct ctttaagacc 780
40 aaaacaaaca agagatggca gtgtaaaaag gaaaaaaaaa actgcgtctg acagagtcgc 840
41 tgaatcaacc atgcttctga taaaaaaaaa aaaaaaaaaa aaaaaaaaa      890
43 <210> SEQ ID NO: 2
44 <211> LENGTH: 224
45 <212> TYPE: PRT
46 <213> ORGANISM: Zea mays
48 <400> SEQUENCE: 2
49 Ser Ala Ala Ala Ala Val Ala Gly Ile Ser Ser Ser Ser Ser Ala Leu
50  1               5                  10                  15
52 Val Thr Ser Thr Val Gly Lys Ser Thr Asn Ile Leu Trp His Glu Cys
53              20                  25                  30
55 Ala Ile Gly Gln Lys Glu Arg Gln Gly Leu Leu Asn Gln Lys Gly Cys
56          35                  40                  45
58 Val Val Trp Ile Thr Gly Leu Ser Gly Ser Gly Lys Ser Thr Leu Ala
```

```
59      50                      55                      60
61 Cys Ala Leu Ser Arg Glu Leu His Gly Arg Gly His Leu Thr Tyr Val
62   65                      70                      75                      80
64 Leu Asp Gly Asp Asn Leu Arg His Gly Leu Asn Arg Asp Leu Ser Phe
65                      85                      90                      95
67 Gly Ala Glu Asp Arg Ala Glu Asn Ile Arg Arg Val Gly Glu Val Ala
68                      100                     105                     110
70 Lys Leu Phe Ala Asp Ala Gly Leu Val Cys Ile Ala Ser Leu Ile Ser
71           115                     120                     125
73 Pro Tyr Arg Ser Asp Arg Ser Ala Cys Arg Asp Leu Leu Pro Lys His
74           130                     135                     140
76 Ser Phe Ile Glu Val Phe Leu Asp Val Pro Leu Gln Val Cys Glu Ala
77 145                     150                     155                     160
79 Arg Asp Pro Lys Gly Leu Tyr Lys Leu Ala Arg Ala Gly Lys Ile Lys
80                      165                     170                     175
82 Gly Phe Thr Gly Ile Asp Asp Pro Tyr Glu Pro Pro Ser Asp Cys Glu
83           180                     185                     190
85 Ile Val Ile Gln Cys Lys Val Gly Asp Cys Pro Ser Pro Glu Ser Met
86           195                     200                     205
88 Ala Gly His Val Val Ser Tyr Leu Glu Thr Asn Gly Phe Leu Gln Asp
89      210                     215                     220
91 <210> SEQ ID NO: 3
92 <211> LENGTH: 1217
93 <212> TYPE: DNA
94 <213> ORGANISM: Zea mays
96 <400> SEQUENCE: 3
97  gcgtccgttt catttcatca atcaaacaga acctctggtc acacacgc agcaaccacc    60
98  gagcccagcg cccggcccag ccagccaggg ccaacggcaa ggcaacaccc tcctcagccc   120
99  gacgccgacg ctcgccgtca tcctcgtaaa tccacagcgc gcgcctcccg tcctcccagg   180
100 cctcacccct agcgatgcgc cactcccggc gctcgtgatc catggcctca ctccccgttc   240
101 ctcacactct tccgcgggtc tcgccagtga tagtgggcgc cgcgaggggg agggccgcgg   300
102 tgcgcgtacg cactgccacc gcggcattgg gcggtgggtg cggcggcggc ggcggaatgg   360
103 agcagcgccc ggggaggccc cgcacagccc agtgaaggag aagcctgtaa tgtcgaacat   420
104 tgggaaatcg actaatattt tatggcacaa ttgcttgatt ggacaatctg atagacagaa   480
105 attgctggga caaaaaggct gtgtcgtatg gataacagga ctcagtggtt cagggaaaag   540
106 tactcttgca tgtgcactga gtcgtgagtt gcattgcaga ggccacctca cgtatgtact   600
107 tgatggtgac aacctcagac atggcctaaa tagagattta agctttaagg cagaagaccg   660
108 tgcagaaaat atacgaagag ttggtgaagt ggcaaagctt tttgctgatg ctggtgtcat   720
109 atgcattgct agcttgatat ctccatacag gagagatcgt gatgcatgcc gtgctctact   780
110 tccacattct aactttattg aagtatttat tgatttgccc ctaaaaattt gtgaagctcg   840
111 tgatcctaaa ggcctataca agcttgcacg tacaggaaag attaaaggtt tcactggaat   900
112 tgatgatcca tacgaaccac caattaatgg tgagatagta attaagatga aagatgagga   960
113 atgcccttca cccaaagcaa tggccaagca agttctatgc taccttgaag aaaacggata  1020
114 tttgcaagct tagtatatgt attttgagaa gattgatctg attcttgtgt gtccattact  1080
115 tgtggacaca ataagatctg ttgttggtca catgaataaa aggcatcaac atgtaggaag  1140
116 taacagaagg tacggttcat tcagaaacgg atatggattc attcgtttaa aaaaaaaaa   1200
117 aaaaaaaaaa aaaaaaa                                                  1217
119 <210> SEQ ID NO: 4
120 <211> LENGTH: 343
```

RAW SEQUENCE LISTING   DATE: 05/08/2002
PATENT APPLICATION: US/09/720,384A   TIME: 17:30:19

Input Set : A:\BB-1167-B USPCT Corrected Seq List 02.txt
Output Set: N:\CRF3\05082002\I720384A.raw

```
121 <212> TYPE: PRT
122 <213> ORGANISM: Zea mays
124 <400> SEQUENCE: 4
125 Arg Pro Phe His Phe Ile Asn Gln Thr Glu Pro Leu Val Thr His Thr
126  1               5                  10                  15
128 Gln Gln Pro Pro Ser Pro Ala Pro Gly Pro Ala Ser Gln Gly Gln Arg
129          20                  25                  30
131 Gln Gly Asn Thr Leu Leu Ser Pro Thr Pro Thr Leu Ala Val Ile Leu
132          35                  40                  45
134 Val Asn Pro Gln Arg Ala Pro Pro Val Leu Pro Gly Leu Thr Pro Ser
135      50                  55                  60
137 Asp Ala Pro Leu Pro Ala Leu Val Ile His Gly Leu Thr Pro Arg Ser
138  65                  70                  75                  80
140 Ser His Ser Ser Ala Gly Leu Ala Ser Asp Ser Gly Arg Arg Glu Gly
141              85                  90                  95
143 Glu Gly Arg Gly Ala Arg Thr His Cys His Arg Gly Ile Gly Arg Trp
144          100                 105                 110
146 Val Arg Arg Arg Arg Arg Asn Gly Ala Ala Pro Gly Glu Ala Pro His
147          115                 120                 125
149 Ser Pro Val Lys Glu Lys Pro Val Met Ser Asn Ile Gly Lys Ser Thr
150      130                 135                 140
152 Asn Ile Leu Trp His Asn Cys Leu Ile Gly Gln Ser Asp Arg Gln Lys
153 145                 150                 155                 160
155 Leu Leu Gly Gln Lys Gly Cys Val Val Trp Ile Thr Gly Leu Ser Gly
156              165                 170                 175
158 Ser Gly Lys Ser Thr Leu Ala Cys Ala Leu Ser Arg Glu Leu His Cys
159          180                 185                 190
161 Arg Gly His Leu Thr Tyr Val Leu Asp Gly Asp Asn Leu Arg His Gly
162          195                 200                 205
164 Leu Asn Arg Asp Leu Ser Phe Lys Ala Glu Asp Arg Ala Glu Asn Ile
165      210                 215                 220
167 Arg Arg Val Gly Glu Val Ala Lys Leu Phe Ala Asp Ala Gly Val Ile
168 225                 230                 235                 240
170 Cys Ile Ala Ser Leu Ile Ser Pro Tyr Arg Arg Asp Arg Asp Ala Cys
171              245                 250                 255
173 Arg Ala Leu Leu Pro His Ser Asn Phe Ile Glu Val Phe Ile Asp Leu
174          260                 265                 270
176 Pro Leu Lys Ile Cys Glu Ala Arg Asp Pro Lys Gly Leu Tyr Lys Leu
177          275                 280                 285
179 Ala Arg Thr Gly Lys Ile Lys Gly Phe Thr Gly Ile Asp Asp Pro Tyr
180      290                 295                 300
182 Glu Pro Pro Ile Asn Gly Glu Ile Val Ile Lys Met Lys Asp Glu Glu
183 305                 310                 315                 320
185 Cys Pro Ser Pro Lys Ala Met Ala Lys Gln Val Leu Cys Tyr Leu Glu
186              325                 330                 335
188 Glu Asn Gly Tyr Leu Gln Ala
189              340
191 <210> SEQ ID NO: 5
192 <211> LENGTH: 431
```

RAW SEQUENCE LISTING   DATE: 05/08/2002
PATENT APPLICATION: US/09/720,384A   TIME: 17:30:19

Input Set : A:\BB-1167-B USPCT Corrected Seq List 02.txt
Output Set: N:\CRF3\05082002\I720384A.raw

```
193 <212> TYPE: DNA
194 <213> ORGANISM: Oryza sativa
196 <220> FEATURE:
197 <221> NAME/KEY: unsure
198 <222> LOCATION: (48)
199 <223> OTHER INFORMATION: n = A, C, G or T
201 <220> FEATURE:
202 <221> NAME/KEY: unsure
203 <222> LOCATION: (346)
204 <223> OTHER INFORMATION: n = A, C, G or T
206 <220> FEATURE:
207 <221> NAME/KEY: unsure
208 <222> LOCATION: (431)
209 <223> OTHER INFORMATION: n = A, C, G or T
211 <400> SEQUENCE: 5
212 cttacacaga gatcaggtag aacagtgggc gagaacaaag ttttgcanat gtcatcaatt  60
213 gtgccgaagg cgtccaatat cttctggcat gattgtgcag ttggccaggc tgatcggcag 120
214 aagctactga agcagaaagg ttgcgttgtt tggatcacag gacttagtgg ttcaggtaaa 180
215 agtaccctgg catgcacatt agatcgagag ctccatacaa gagggaagct ttcttatgtt 240
216 cttgatggtg ataatttaag acatggtttg aacaaggatc ttggctttaa ggcggaagac 300
217 cgtgctgaaa atatacgcaa agttggtgag gtagcaaagc tattcncaga tgcaagccta 360
218 gtatgcattg caagtttcaa atctccctat aagagagaac gtgagtcctg gccctgcaat 420
219 attgtcaaat n                                                      431
221 <210> SEQ ID NO: 6
222 <211> LENGTH: 118
223 <212> TYPE: PRT
224 <213> ORGANISM: Oryza sativa
226 <220> FEATURE:
227 <221> NAME/KEY: UNSURE
228 <222> LOCATION: (98)
229 <223> OTHER INFORMATION: Xaa = ANY AMINO ACID
232 <400> SEQUENCE: 6
233 Ser Ile Val Pro Lys Ala Ser Asn Ile Phe Trp His Asp Cys Ala Val
234  1               5                  10                  15
236 Gly Gln Ala Asp Arg Gln Lys Leu Leu Lys Gln Lys Gly Cys Val Val
237             20                  25                  30
239 Trp Ile Thr Gly Leu Ser Gly Ser Gly Lys Ser Thr Leu Ala Cys Thr
240             35                  40                  45
242 Leu Asp Arg Glu Leu His Thr Arg Gly Lys Leu Ser Tyr Val Leu Asp
243         50                  55                  60
245 Gly Asp Asn Leu Arg His Gly Leu Asn Lys Asp Leu Gly Phe Lys Ala
246  65                  70                  75                  80
248 Glu Asp Arg Ala Glu Asn Ile Arg Lys Val Gly Glu Val Ala Lys Leu
249             85                  90                  95
251 Phe Xaa Asp Ala Ser Leu Val Cys Ile Ala Ser Phe Lys Ser Pro Tyr
252             100                 105                 110
254 Lys Arg Glu Arg Glu Ser
255         115
257 <210> SEQ ID NO: 7
```

```
258 <211> LENGTH: 936
259 <212> TYPE: DNA
260 <213> ORGANISM: Glycine max
262 <400> SEQUENCE: 7
263 gcacgagcca ccgcgaaggc tctgcgacag ccctgctacg ccggaatctt tcgcaacatc  60
264 gaatgcggcc cgtcgccggc ggcggagtcg ctagggtttc cgaagctccg cggaatcaac 120
265 gtcactggat tgcactgcgg ccgccgaggc ctcgtcctcg tcctccgtgc aaaatcaaag 180
266 ccgattaggg cgaaggagaa cgcaagcgta agtgcttctc tgatcgatga ctggttcaag 240
267 ccaattacgg cgaaggagga ttctaacgca gaggaccgta catcttcgtt ttctggtaaa 300
268 aatctcaccc agatgtcaaa tgttgggaac tcgacaaaca ttatgtggca tgactgtcca 360
269 attcagaaac aagatagaca gcagctgctt cagcaacaag gctgtgttat atggctaact 420
270 ggcctcagcg gatcaggaaa aagcactatt gcatgtgctc tgagtcaaag cttgcactcc 480
271 aaaggaaaac tgtcttacat ccttgatggt gacaatattc ggcatggtct aaaccaggat 540
272 cttagtttta gagcagaaga tcgttctgaa aacattagaa ggattggtga ggtggcaaaa 600
273 ctctttgcag atgctggtgt tatttgcatc actagtttaa tatcaccata ccaaaaggat 660
274 agagatgcat gcagagcact actttcaaaa ggagatttta ttgaggtttt catagatgtt 720
275 ccactacatg tgtgtgaagc tagggaccca aagggactct acaagcttgc tcgagctgga 780
276 aagatcaaag gtttcactgg tatagatgat ccatatgaac caccgtgtag ttgtgagata 840
277 gtattacaac agaaaggaag tgactgtaag tctcccagtg atatggctga agaagtgata 900
278 tcctacttgg aggagaacgg atacctgcgg gcttga                            936
280 <210> SEQ ID NO: 8
281 <211> LENGTH: 311
282 <212> TYPE: PRT
283 <213> ORGANISM: Glycine max
285 <400> SEQUENCE: 8
286 Ala Arg Ala Thr Ala Lys Ala Leu Arg Gln Pro Cys Tyr Ala Gly Ile
287  1               5                  10                  15
289 Phe Arg Asn Ile Glu Cys Gly Pro Ser Pro Ala Ala Glu Ser Leu Gly
290              20                  25                  30
292 Phe Pro Lys Leu Arg Gly Ile Asn Val Thr Gly Leu His Cys Gly Arg
293          35                  40                  45
295 Arg Gly Leu Val Leu Val Leu Arg Ala Lys Ser Lys Pro Ile Arg Ala
296      50                  55                  60
298 Lys Glu Asn Ala Ser Val Ser Ala Ser Leu Ile Asp Asp Trp Phe Lys
299  65                  70                  75                  80
301 Pro Ile Thr Ala Lys Glu Asp Ser Asn Ala Glu Asp Arg Thr Ser Ser
302              85                  90                  95
304 Phe Ser Gly Lys Asn Leu Thr Gln Met Ser Asn Val Gly Asn Ser Thr
305              100                 105                 110
307 Asn Ile Met Trp His Asp Cys Pro Ile Gln Lys Gln Asp Arg Gln Gln
308              115                 120                 125
310 Leu Leu Gln Gln Gln Gly Cys Val Ile Trp Leu Thr Gly Leu Ser Gly
311          130                 135                 140
313 Ser Gly Lys Ser Thr Ile Ala Cys Ala Leu Ser Gln Ser Leu His Ser
314 145                 150                 155                 160
316 Lys Gly Lys Leu Ser Tyr Ile Leu Asp Gly Asp Asn Ile Arg His Gly
317              165                 170                 175
319 Leu Asn Gln Asp Leu Ser Phe Arg Ala Glu Asp Arg Ser Glu Asn Ile
320              180                 185                 190
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 05/08/2002
PATENT APPLICATION: US/09/720,384A   TIME: 17:30:20

Input Set : A:\BB-1167-B USPCT Corrected Seq List 02.txt
Output Set: N:\CRF3\05082002\I720384A.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:5; N Pos. 48,346,431
Seq#:6; Xaa Pos. 98